

GHRC Cloud

GHRC Cloud Update

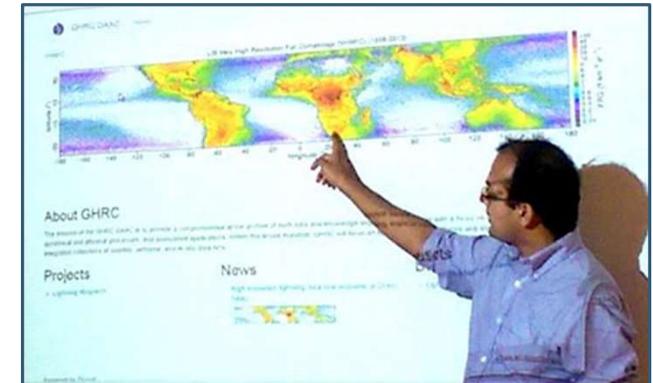
Will Ellett
GHRC Operations Manager



Team Roles & Responsibilities

Project Manager
DAAC Scientist
System Architect
System Engineer
Facilitator
LIS Team Lead
Data Scientist
Developer
Operator
IT (SyAdmin)

Will Ellett
Geoffrey Stano
Ajinkya Kulkarni
Abdelhak Marouane
Taylor Wright
Sherry Harrison
Leigh Sinclair, Lucy Wany, Xiang Li
Amy, Lin, Brian Ellingson, Eddie Campos
Lamar Hawkins, Shannon Flynn, Mary Nair
Michele Garrett, Michael McEniry



GHRC Cloud Migration team received 2020 NASA Group Achievement Award

“For commitment to innovation, collaboration, and teamwork in the transition of the NASA Global Hydrology Resource Center data center to the cloud.”

1

Trained Operators/DMG

Operators and Data Management group trained on how to publish datasets in the cloud

2

Began Parallel Operations

During October 2019, GHRC started publishing all datasets on-prem and in the cloud

3

Migrated to AWS West

GHRC moved its data and operation from AWS East to AWS West to align with other ESDIS cloud activities

4

Validated Cloud Datasets

After the move to AWS West, GHRC validated all cloud static datasets

5

Updated HyDRO

Once static cloud datasets were validated, HyDRO was updated to point to cloud URL's

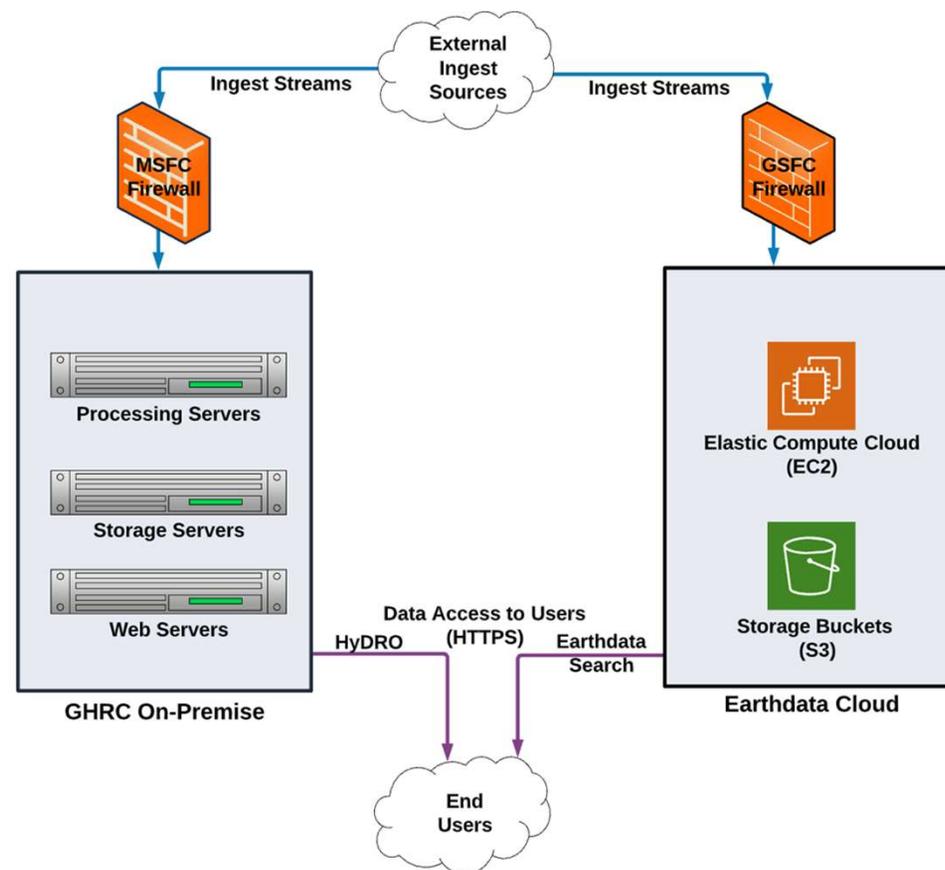
6

Passed Readiness Reviews

Successfully passes the ESDIS Review Gate & Production Readiness Review for Parallel Operations

Parallel Operations

- GHRC is currently operating on-premise and in Earthdata Cloud
- All new GHRC datasets are ingested in both environments
- Verifying Cloud Near Real-Time and On-Going holdings match on-prem using tool developed by GHRC
- Continuing to refine and improve Near Real-Time (NRT) and On-Going workflows

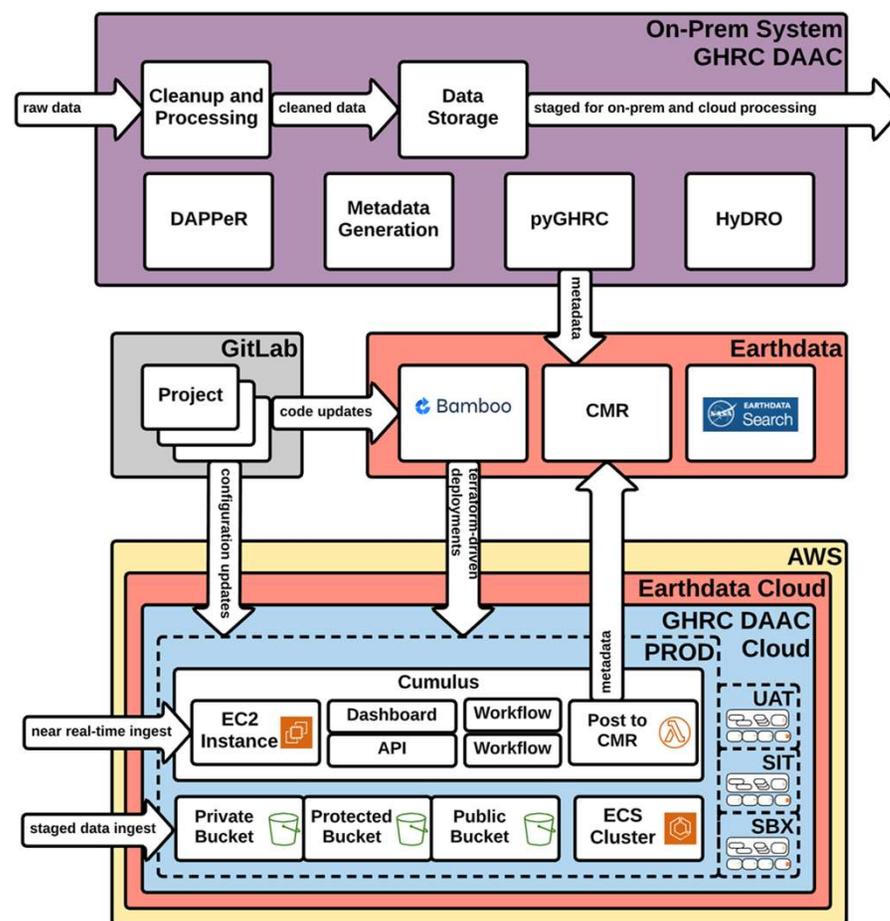


System Architectural Overview

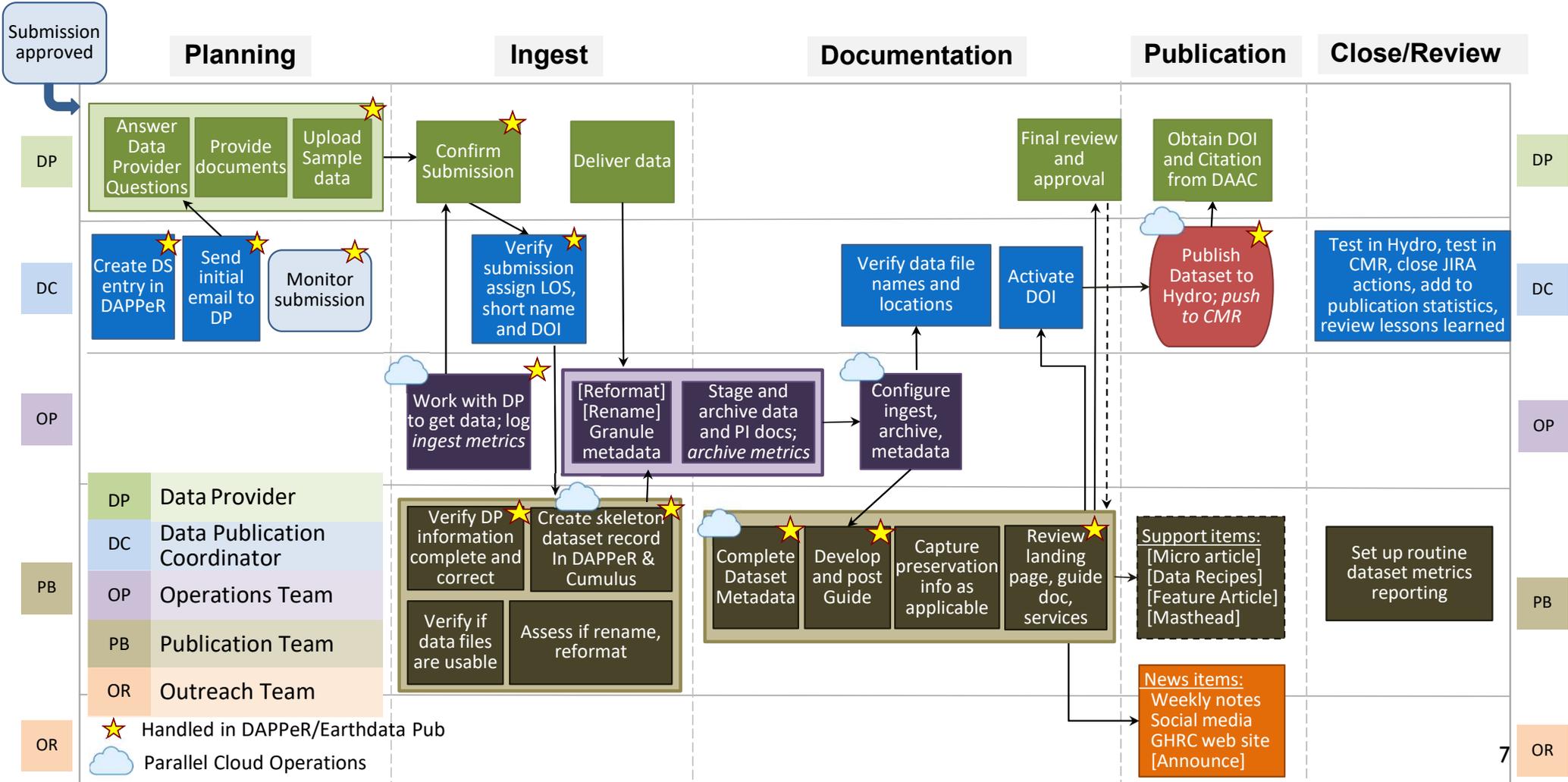
Data Ingest: The core of the ingest and processing capabilities in Cumulus is built into the deployed AWS Step Function and EC2 workflows.

Data Archive: GHRC data is stored in AWS S3 and archived in non-NGAP Glacier storage until the Operational Recovery Cloud Archive (ORCA) is ready.

Data Distribution: End users can access the GHRC data using the Cumulus distribution module which provides Earthdata Login authenticated data egress, temporary S3 links as well as metrics.



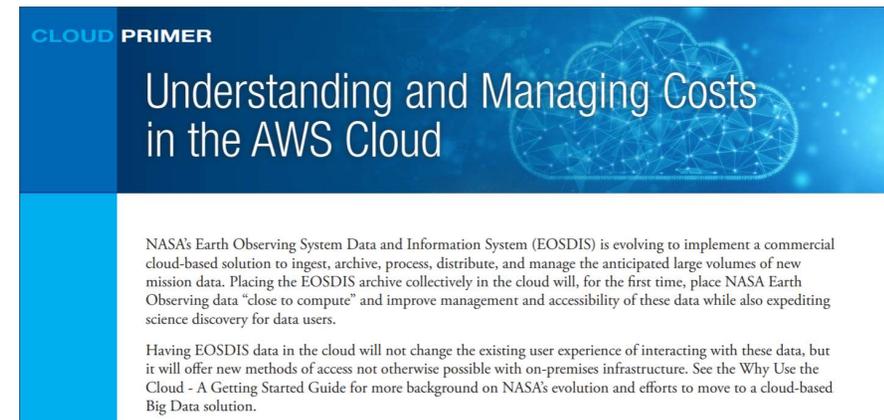
GHRC Data Publication Process



GHRC has contributed several improvements to Cumulus and other cloud tools

Examples:

- Provided Cumulus code fixes and recommendations
- Participated in creation of Cloud Primer
- Pathfinder for operating DAAC in the cloud
 - Created DMR++ for OPeNDAP
 - Created python Cumulus API command-line utilities (now used by several DAACs)
 - Helped stabilize the base code for an Operational Cloud Backup



CLOUD PRIMER

Understanding and Managing Costs in the AWS Cloud

NASA's Earth Observing System Data and Information System (EOSDIS) is evolving to implement a commercial cloud-based solution to ingest, archive, process, distribute, and manage the anticipated large volumes of new mission data. Placing the EOSDIS archive collectively in the cloud will, for the first time, place NASA Earth Observing data "close to compute" and improve management and accessibility of these data while also expediting science discovery for data users.

Having EOSDIS data in the cloud will not change the existing user experience of interacting with these data, but it will offer new methods of access not otherwise possible with on-premises infrastructure. See the Why Use the Cloud - A Getting Started Guide for more background on NASA's evolution and efforts to move to a cloud-based Big Data solution.

<https://earthdata.nasa.gov/learn/user-resources/webinars-and-tutorials/cloud-primer>

Cloud-only Requirements

1

DAPPeR & HyDRO Replacement

Need Earthdata Pub and Earthdata Search Client support

2

Sustainable Cloud Ops

Need improved troubleshooting and alert notification capabilities

3

Operational Backup

Need cloud-based backup with support for recovery as well as tertiary off-site backup

4

Operator Work Area

Need area for operators to perform data manipulations before publication

5

Upload Area Support

Need to allow data providers a location to upload field campaign data

6

Website Migration

Need to migrate GHRC website content to the cloud

Upcoming Activities

Next 6 months

- Update HyDRO to point near real-time and on-going datasets to cloud links
- Convert data recipes to interactive notebooks
- Prototype Operator area in cloud
- Distribute ISS LIS NRT data via cloud
- Add browse images to cloud workflows
- Continue Cloud Pathfinder Efforts
 - Improve age-off capabilities
 - Improve watchdog notifications
 - Improve troubleshooting capabilities

Next 12 months

- Migrate GHRC web capabilities to Earthdata Cloud
- Continue to recommend improvements to the Earthdata Cloud team
- Investigate making selected GHRC datasets available in cloud-optimized data format
- Investigate prototyping of services in Harmony to make FCX data available through APIs
- Continue to streamline Data Publication process
- Prepare for Cloud-only operations

GHRC Cloud

THANK YOU!

QUESTIONS?

